



OPEN

A perceptual glitch in serial perception generates temporal distortions

Franklenin Sierra^{1✉}, R. Muralikrishnan¹, David Poeppel^{2,3} & Alessandro Tavano¹

Precisely estimating event timing is essential for survival, yet temporal distortions are ubiquitous in our daily sensory experience. Here, we tested whether the relative position, duration, and distance in time of two sequentially-organized events—standard S, with constant duration, and comparison C, with duration varying trial-by-trial—are causal factors in generating temporal distortions. We found that temporal distortions emerge when the first event is shorter than the second event. Importantly, a significant interaction suggests that a longer inter-stimulus interval (ISI) helps to counteract such serial distortion effect only when the constant S is in the first position, but not if the unpredictable C is in the first position. These results imply the existence of a perceptual bias in perceiving ordered event durations, mechanistically contributing to distortion in time perception. We simulated our behavioral results with a Bayesian model and replicated the finding that participants disproportionately expand first-position dynamic (unpredictable) short events. Our results clarify the mechanisms generating time distortions by identifying a hitherto unknown duration-dependent encoding inefficiency in human serial temporal perception, something akin to a strong prior that can be overridden for highly predictable sensory events but unfolds for unpredictable ones.

Precisely estimating event timing is essential for a range of perceptual and cognitive tasks, yet temporal distortions are ubiquitous in our daily sensory experience^{1–3}. A specific kind of time distortion is the presentation-order error⁴. In 1860, Fechner observed that when comparing the weight of two elements, the order in which they were lifted mattered⁵. This led to a systematic error on a subject's judgment of sequentially presented stimuli, which was termed time-order error (TOE)⁴. Time-order errors have been detected in different stimulus modalities, such as audition, vision, and taste, as well as different stimulus dimensions, such as loudness, heaviness, and brightness⁴. Understanding how they are generated is fundamental as humans ordinarily perceive events in a series, not in isolation.

Time-order errors in temporal judgment can be experimentally tested by implementing a two-interval forced choice (2IFC) discrimination task, whereby participants compare the duration of two successive time intervals (events) per trial—a Standard and a Comparison (S vs C)—separated by an inter-stimulus interval (ISI)^{6,7}. When combined with the method of constant stimuli, the duration of S is kept fixed across the experimental session, whereas the duration of C changes from trial to trial and usually it can take one of six to nine durations distributed around the S duration⁸.

In a 2IFC task, temporal performance is modelled by fitting a psychometric function. From this fitting, two main dependent variables are obtained: the point μ where the curve cuts the 50% line (that is, the point of subjective equality [PSE]) and the slope of the resulting curve. While the PSE estimates the accuracy of the comparison judgment—and provides a marker for temporal distortions—the slope estimates their temporal precision^{9,10}. Time order error effects have recently classified into two types: effects of the stimulus order on the PSE are called Type A effects, whereas effects on temporal precision are called Type B effects^{11,12}. It is important to note that there exists another type of mistake called the contraction bias: When the 1st stimulus is small relatively to the distribution of the stimuli used in the experiment, participants tend to overestimate it, whereas when it is large relatively to the distribution of the stimuli used in the experiment, they tend to underestimate it^{13,14}.

Traditionally, TOEs have been variously attributed to sensory desensitization^{7,15}, poor sensory weighting of C relative to S^{16,17}, or idiosyncratic response bias¹⁵. However, more recently, two additional models have attempted to explain TOE: (1) the internal reference model (IRM)¹⁸—an updated version of the sensory desensitization model—and (2) Bayesian observer models^{19–21}.

¹Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, HE, Germany. ²Department of Psychology, New York University, New York City, NY, USA. ³Ernst Strüngmann Institute for Neuroscience, Frankfurt am Main, HE, Germany. ✉email: franklenin.sierra@gmail.com

The underlying idea behind IRM is that in comparing S and C, participants maintain an internal representation with the average duration of previous trials. However, this internal representation is updated by taking into consideration only the first presented stimulus (S or C). Because the S stimulus has a constant duration and the C stimulus varies unpredictably, more errors will be made when the order of presentation is <CS> compared to when it is <SC>. Raviv et al.¹⁴ proposed a Bayesian observer model, which is akin to the IRM but with a Bayesian inference approach. Such a model assumed that the brain uses a heuristic strategy to discriminate auditory temporal intervals, with the idea being that when a human participant compares two stimuli, the second auditory interval is compared against the decaying average of the first one. Raviv et al., suggested that errors in temporal discrimination arise during memory retrieval/decision making and not during memory encoding.

Bayesian models offer a dynamical approach and take into consideration the representation of the two stimuli, S and C. Thus, the result of perceiving an S or C interval (modeled as a posterior distribution) is the product of the previous trial's representation—and all the information collected until that point—(modeled as a prior distribution) and the current sensory input (the likelihood function). In this model, the prior distribution is updated from trial to trial, whereas the posterior distribution is modulated at each trial by the perception of both stimuli, S and C. A common strategy used to model the update of the prior and posterior distributions is by using a Kalman filter^{22,23}. Indeed, de Jong et al.¹² found that in comparing the duration of visual stimuli, the influence of statistical context on time estimation is best explained by a Bayesian model using a Kalman filter, and thus discarded the IRM model. The authors found that Type A effects are influenced by a dynamic prior that is sequentially updated by both stimuli, S and C.

In our previous work, we showed that in discriminating two successive visual events (each event signaled by two blue disks) with $S < 200$ ms, time distortions appear only if the ISI is shorter than 1 s²⁴. Here, we focused on the type A effect and tested how the factors that determine serial dynamics of relative event duration—relative position, relative distance in time, and relative duration of S and C—contribute to generating temporal distortions. We used an S of 120 ms and varied the ISI over four different intervals (400, 800, 1600, and 2000 ms).

Firstly, we swapped the order of presentation of S and C (relative position factor). Secondly, we tested whether a long ISI (relative distance factor) increases the temporal accuracy. Finally, we tested whether the ordinal position of the longer stimulus (relative duration factor) modulates temporal accuracy, under the assumption that the contraction bias may apply locally, that is for relative durations within a trial, and independently of event type (S or C). For this, we used equiprobable C durations distributed around the duration of the S stimulus. We thus hypothesized that: (1) with an ever-changing C in first position, time distortions would increase as participants would benefit to a markedly lesser extent from increased attention orienting for long ISIs. Hence, we expected an interaction between the two factors: stimulus presentation order and ISI; (2) if distortions in duration comparison are mainly resultant from the predictability features of S and C (trial-by-trial predictable vs. unpredictable), then the ordinal position of the longer stimulus should not modulate temporal perception.

Results verify hypothesis 1: the best Generalized Linear Mixed Model (GLMM) included an interaction between stimulus presentation order (<SC> or <CS>) and ISI. Increasing ISI reduces temporal distortions, more so for the <SC> group. Surprisingly, however, and contrary to hypothesis 2, the relative position of the longer stimulus has important modulatory effects on temporal perception. Not all first-position events are subjectively expanded to the extent that they produce distortions in temporal judgment. Instead, time distortions tend to be generated when the first event in a series is shorter than the second event, independent of event type (S or C). Notably though, when the dynamic stimulus (C) is in first position, the ensuing distortion effect cannot be compensated for by increasing the ISI.

To dig deeper into the mechanics of TOEs and show the computational plausibility of the highlighted perceptual bias, we considered the fixed factors of the best GLMM and simulated our findings by implementing a Bayesian model using a Kalman filter^{12,22,23}. Model results confirm the findings on human participants: first-position shorter events are disproportionately expanded. Our results contribute to clarifying the mechanics generating perceptual time distortions, by identifying a novel duration-dependent encoding inefficiency in human serial time perception.

Results

Two separate groups of human participants performed a 2IFC discrimination task comparing the duration of an S event against that of a C event (or vice versa), and deciding which stimulus was longer. To signal the onset and offset of each event, we used a short-duration blue disk (hence, S and C were empty visual stimuli, see “[Material and methods](#)” section). For the <SC> experimental group, the S stimulus was displayed in the first position, and was shifted to the second position for the <CS> experimental group (Fig. 1a). The duration of the S event was kept constant (120 ms), whereas the duration of the C event varied, providing participants with three degrees of sensory evidence (weak $\pm \Delta$ 20 ms, medium $\pm \Delta$ 60 ms, and strong $\pm \Delta$ 100 ms; Fig. 1b). We parametrically manipulated the ISI by using four durations: 400, 800, 1600, and 2000 ms.

Temporal accuracy. Temporal accuracy was 84.39% (SD = 7.16) and 79.75% (SD = 9.79) for the <SC> and <CS> groups, respectively. Accuracy improved across the board for ISI conditions > 400 ms, however accuracy values were always lower for the <CS> than for the <SC> group (Fig. 1c; Table 1).

Raw accuracy was analyzed with GLMMs using a binomial parameter with a logit link function^{10,25}. Model selection was based on the Akaike Information Criterion (AIC), whereby the model with the lowest AIC was selected as the one that best explained the data (see “[Material and methods](#)”). Results showed that the best model for explaining the data included all the factors: Δ , ISI, Group, and Ordinal position of the longer stimulus, and their interaction term, with random intercepts for each participant (AIC = 39,322.53). Models involving less complex / no random intercept specifications either did not converge or had higher AIC.

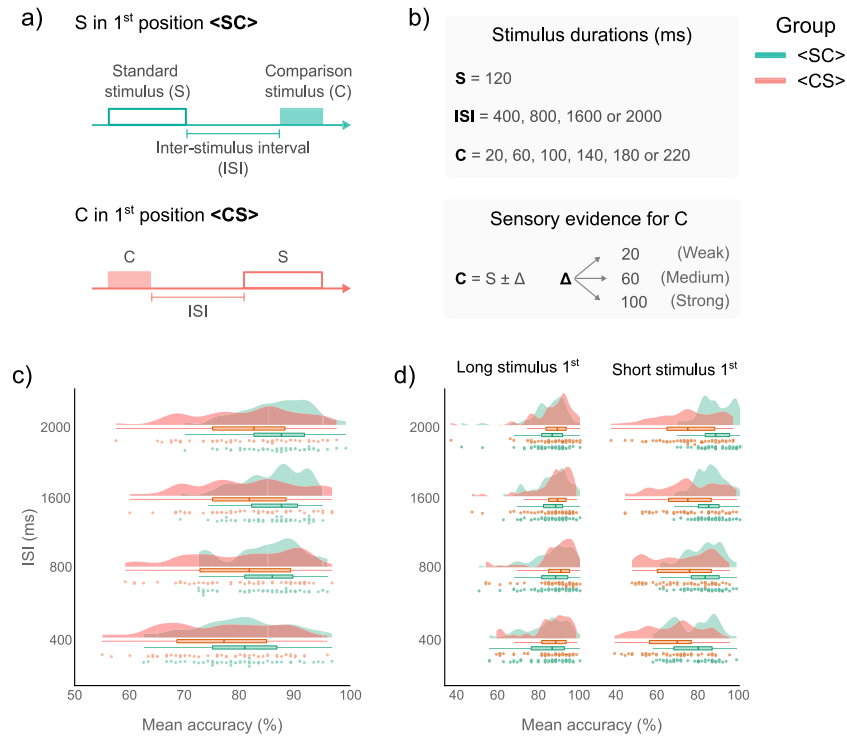


Figure 1. Two-interval forced choice (2IFC) task and accuracy. **(a)** Timeline of events in the 2IFC task. For the <SC> group the standard stimulus (S) was displayed in the 1st position and in the 2nd position for the <CS> group. **(b)** The S stimulus had a fixed duration of 120 ms, whereas the comparison stimulus (c) varied trial-by-trial according to its level of sensory evidence: weak, medium, or strong. We implemented four Inter-stimulus intervals (ISIs: 400, 800, 1600 and 2000 ms). **(c)** Mean accuracy for each group and each ISI level. Data points depict the mean accuracy for each participant. Box and density plots show the distribution of the mean accuracy for both <SC> and <CS> groups. The median is represented by the vertical line in the box plots, whereas the right and left edges depict the interquartile range (IQR). Accuracy is higher for the <SC> group, however for both groups the accuracy increased for ISI > 400 ms. **(d)** Mean accuracy separated by the ordinal position of the longer stimulus. Accuracy for the <CS> group decreases when the 1st stimulus is shorter than the 2nd stimulus.

Group	ISI level			
	400 ms	800 ms	1600 ms	2000 ms
<SC>	80.52% (8.30)	84.59% (6.62)	85.89% (5.80)	86.55% (6.27)
<CS>	76.41% (10.03)	80.21% (9.89)	81.18% (9.25)	81.20% (9.45)

Table 1. Accuracy. Mean accuracy and the standard deviation (SD, in parentheses) for each group and each ISI level. In both groups, participants had their worst performance at the ISI₄₀₀ condition. However, mean accuracy values are lower for the <CS> group than for the <SC> group.

Type II Wald chisquare tests on the selected model showed a main effect of Delta ($\chi^2(2) = 3467.69, p < 0.001$), ISI ($\chi^2(3) = 190.40, p < 0.001$), Group ($\chi^2(1) = 6.60, p < 0.001$), and Ordinal position of the longer stimulus ($\chi^2(1) = 545.93, p < 0.001$), their interactions were also significant (Table 2). As the best model included the Ordinal position of the longer stimulus as fixed effect (and their interactions), this suggest that the position of the longer stimulus, regardless of their experimental function—Standard or Comparison—modulates accuracy (Fig. 1d).

To examine the effects of the Ordinal position of the longer stimulus, we deployed contrast analyses on the estimated marginal means (EMMs) of the best GLMM, with bonferroni adjustment of *p*-values for multiple comparisons where applicable. Results showed that for both groups <SC> and <CS> the Ordinal position of the longer stimulus has modulatory effects on the accuracy ($p = 0.0254; p < 0.0001$; respectively). However, when the long stimulus is displayed in 1st position there are not differences between Groups, the opposite occurs when the long stimulus is displayed in 2nd position ($p = 0.8363; p < 0.0001$; respectively; Fig. 1d).

Fixed factors and interactions	χ^2	df	p. value
Delta	3467.69	2	<0.001***
ISI	190.40	3	<0.001***
Long stimulus first	545.93	1	<0.001***
Group	6.60	1	0.010*
Delta:ISI	78.22	6	<0.001***
Delta:Long stimulus first	295.46	2	<0.001***
ISI:Long stimulus first	75.89	3	<0.001***
Delta:Group	46.77	2	<0.001***
ISI:Group	13.19	3	<0.010**
Long stimulus first:Group	325.33	1	<0.001***
Delta: ISI:Long stimulus first	30.94	6	<0.001***
Delta:ISI:Group	11.70	6	0.068
Delta:Long stimulus first	168.20	2	<0.001***
ISI:Long stimulus first:Group	10.25	3	<0.050*
Delta:ISI:Long stimulus first : Group	19.73	6	<0.010**

Table 2. Statistical table. Results of the II Wald χ^2 test on the best model explaining accuracy. All fixed factors and their interaction term, except for the Delta \times ISI \times Group interaction, were statistically significant. (Significant codes: ***, p -value < 0.001; **, a p -value < 0.01; and *, a p -value < 0.05).

Group	ISI	Estimate	SE	df	z. ratio	p. value
<SC>	400	0.29	0.07	Inf	3.87	<0.001***
	800	0.18	0.09	Inf	1.91	0.055
	1600	0.07	0.10	Inf	0.72	0.467
	2000	-0.13	0.09	Inf	-1.42	0.154
<CS>	400	0.84	0.07	Inf	11.98	<0.001***
	800	0.74	0.08	Inf	8.97	<0.001***
	1600	0.45	0.08	Inf	5.40	<0.001***
	2000	0.54	0.08	Inf	6.28	<0.001***

Table 3. Pairwise comparisons of estimated marginal means (EMMs) for each ISI level. Post-hoc comparisons on the Ordinal position of the longer stimulus, within each Group and each level of ISI, revealed that for the <SC> group the only statistically significant comparison was at the ISI₄₀₀, whereas for the <CS> group all pairwise comparisons were statistically significant. (***) p -value < 0.001).

Group	Delta	Estimate	SE	df	z. ratio	p. value
<SC>	20	0.18	0.04	Inf	3.74	<0.001***
	60	0.05	0.07	Inf	0.69	0.485
	100	0.07	0.10	Inf	0.69	0.485
<CS>	20	1.54	0.05	Inf	30.19	<0.001***
	40	1.04	0.06	Inf	17.13	<0.001***
	60	-0.65	0.09	Inf	-7.01	<0.001***

Table 4. Pairwise comparisons of EMMs for each Delta level. Post-hoc comparisons on the Ordinal position of the longer stimulus, within each Group and each Delta level, revealed that for the <SC> group the only statistically significant comparison was at the Δ 20, whereas for the <CS> group all pairwise comparisons were statistically significant. (***) p -value < 0.001).

Pairwise comparisons at the ISI levels of each group revealed statistically significant differences between EMMs only at the ISI₄₀₀ level for the <SC> group, whereas for the <CS> group we found statistically significant differences at each ISI level (all p s < 0.0001; Table 3), suggesting that an increasing ISI does not help suppressing distortions when the first stimulus does not have a constant duration.

Results of contrast analyses at each Δ levels showed that for the <CS> group all pairwise comparisons were statistically significant (all p s < 0.0001; Table 4). However, for the <SC> group results revealed a significant difference only at the Δ 20 level (p < 0.0002; Table 4). These results suggest the existence of a perceptual bias that

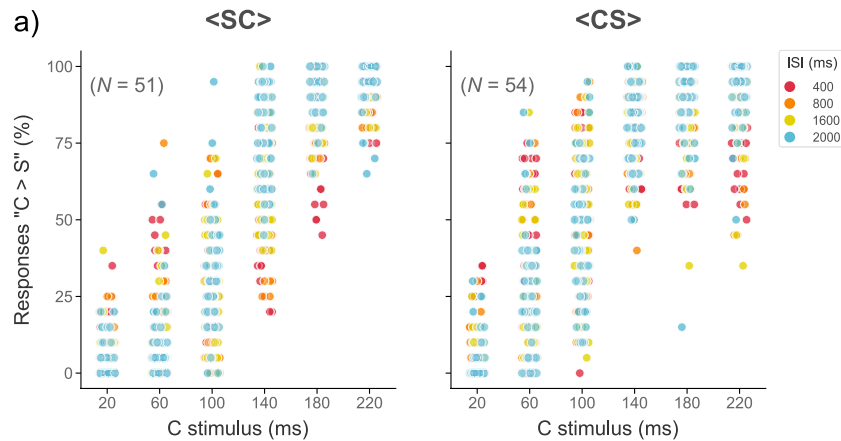


Figure 2. “C > S” responses. **(a)** Percentage of “C > S” responses plotted as a function of the C stimulus. Data points depict individual mean responses at each ISI level. For the <SC> group participants tend to make more mistakes at the 140-ms C stimulus, whereas for the <CS> group they tend to make more mistakes at the 100-ms C stimulus. We analyzed these behavioral responses with generalized linear mixed models (GLMMs) to estimate a single model across all subjects.

can be minimized by both an increase of the ISI or providing more sensory evidence (that is, Δ), but only if the first event has a constant duration.

Constant error (CE). To obtain the temporal sensitivity and the magnitude of the time distortions (indexed via the PSE) of each group, we fitted GLMMs using the percentage of responses “C longer than S” (Fig. 2a). We used as predictors the fixed factors of the previous model, except for the “Ordinal position of the longer stimulus” factor. Here again, we used a binomial parameter.

Based on the AIC the best model included the C stimulus, ISI, Group, and their interaction term as fixed effects, with random intercepts for each participant and random slopes for the effect of the ISI for each participant (AIC = 40,208.74; Fig. 3a). Type II Wald chisquare tests on the selected model showed a main effect of C stimulus and Group ($\chi^2(5) = 18457.67, p < 0.001$; $\chi^2(1) = 64.53, p < 0.001$; respectively). All interactions (C stimulus \times ISI; C stimulus \times Group; ISI \times Group; C stimulus \times ISI \times Group) were also significant (all $ps < 0.001$; Table 5).

To obtain the individual temporal sensitivity and the PSE, we fitted a GLM for each participant by using as predictors the fixed effects of the best GLMM. We derived the CE from the PSE to obtain the exact magnitude of the time distortions (see “Material and methods” section). Results showed that, group-wise, CE values decrease with increasing ISI regardless of the group. However, CE values were higher for the <CS> than for the <SC> group (Fig. 3b; Table 6). Indeed, a Mixed Bayesian ANOVA revealed that best model for explaining these data was the model including the factors ISI and Group ($BF_{10} = 5.7 \times 10^7$).

Bayesian model. We implemented a Bayesian model to replicate our results by using a Kalman filter^{12,22,23}. For each group we simulated data of 100 subjects using 120 trials at each ISI level (see “Material and methods” section). We applied the best GLMM of the human observers to the simulated data. As with the human participants, we applied a GLM to each subject to obtain the individual temporal sensitivity and the PSE. To compare the responses of the human observers against the Bayesian observer’s responses, we obtained the root mean squared error (RMSEs). Results showed that the Bayesian observer’s responses successfully simulated the trend of results of the human observers: (1) the CEs decrease with an increase of the ISI; (2) CEs values are higher for the <CS> than for the <SC> group (Fig. 4a, b).

Discussion

The duration of an event can be distorted when the event is inserted in a series. Such distortion, termed time-order error (TOE), constitutes one of the oldest and most investigated phenomena of subjective time perception^{5,17,26}. Yet, the mechanics of TOE generation are still unclear. Since TOEs occur in serial discrimination tasks, we tested how event duration dynamics—relative position, distance in time and duration of two successive events—contributes to time distortions by flipping the positions of S and C events in two separate behavioral experiments.

We report on four main findings. First, in line with previous work, the interaction between stimulus order presentation and ISI suggest that, with an ever-changing and therefore unpredictable stimulus in first position, temporal accuracy decreases. These differences in accuracy contribute to the Type A effect: smaller CEs for the <SC> group and larger CEs for the <CS> group. Dyjas et al.¹⁸ found no significant statistical differences for the Type A effect between <SC> and <CS> presentation orders, as far as both visual and auditory modalities are concerned. On the contrary, our findings show that at least when using empty visual events, the serial order of presentation <SC> and <CS> modulates temporal accuracy and thus the Type A effect arise. Second, we

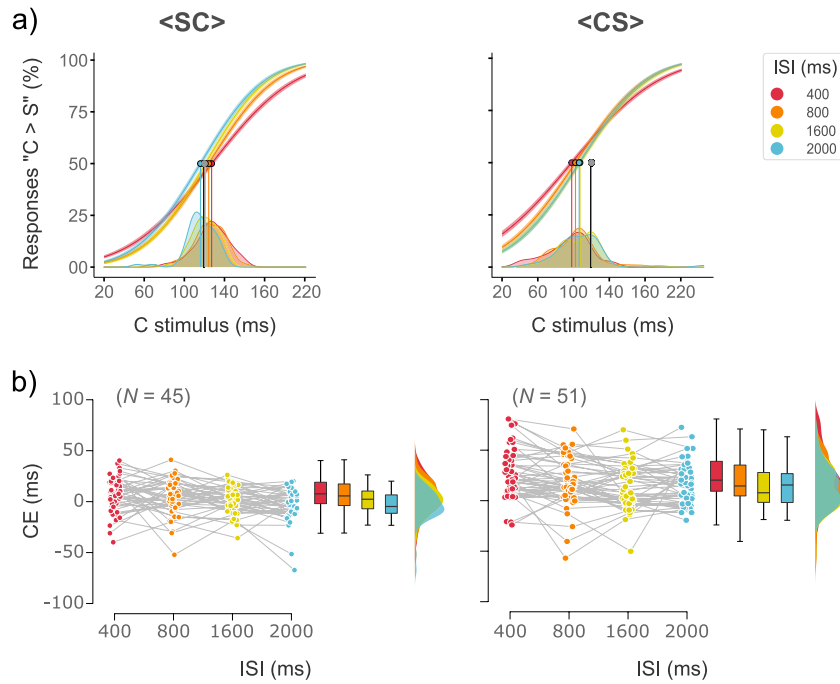


Figure 3. Constant error (CE) of human observers. **(a)** Psychometric curves of human observers. Fitted curves modeling performance of each ISI level for both groups: <SC> and <CS>. The durations of the six C stimuli are plotted on the x-axis and the probability of responding “C longer than S” on the y-axis. Logistic curves depict separate fits for each ISI condition. The gray dot and vertical line depicts the physical magnitude ϕ_s of the S stimulus. The remaining dots and vertical lines represent the point of subjective equality (PSE) for each ISI condition. Density plots show the subject-to-subject variability of the PSE for each ISI level. **(b)** Data points depict individual CEs for both <SC> and <CS> groups at each ISI level. Box-whisker, and density plots show the distribution of the CEs. The median is represented by the horizontal line in the box plots, whereas the bottom and top whiskers depict the IQR. CEs are larger for the <CS> than for the <SC> group. However, for both groups the CE decreases as the ISI increases, with exception of the ISI_{2000} condition of the <CS> group. Note that subjects with 3 standard deviations above/below the mean of the CE and slope were discarded.

Fixed factors and interactions	χ^2	df	p. value
C stimulus	18,457.67	5	< 0.001***
ISI	3.87	3	0.275
Group	64.53	1	< 0.001***
C stimulus:ISI	293.18	15	< 0.001***
C stimulus:Group	480.01	5	< 0.001***
ISI:Group	18.28	3	0.003***
C stimulus:ISI:Group	63.51	15	< 0.001***

Table 5. Statistical table. Results of the II Wald χ^2 test on the best model explaining “C > S” responses. The C stimulus and the Group factors with their interaction term were significant.

Group	ISI level			
	400 ms	800 ms	1600 ms	2000 ms
<SC>	6.78 (17.45)	4.52 (17.94)	1.27 (13.11)	- 4.23 (16.26)
<CS>	25.46 (24.19)	17.86 (23.46)	13.45 (21.82)	15.35 (20.24)

Table 6. Constant errors (CEs). Mean CE values and SDs for each group and each ISI level. CEs were larger for the <CS> than for the <SC> group. However, for both groups the CE decreases as the ISI increases, with exception of the ISI_{2000} condition of the <CS> group.

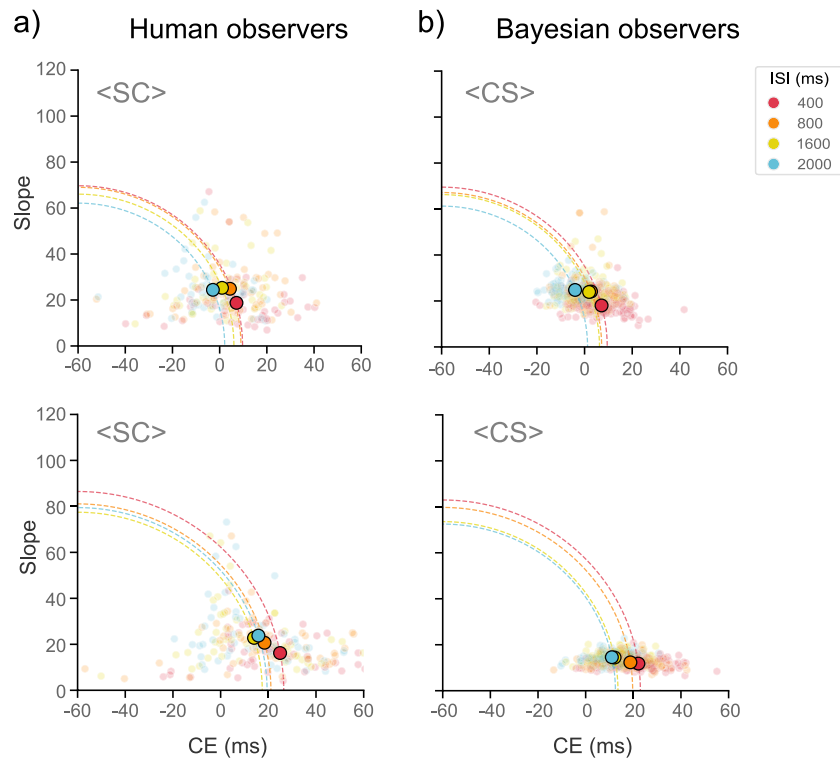


Figure 4. Root mean squared errors (RMSEs) of human and Bayesian observers. **(a)** RMSEs of human observers. RMSEs are given by the distance from the origin and are depicted by a quarter circle. Any increase in the CE or the slope will lead to a larger radius. Big dots depict the intersection of the CE and the slope's mean for each group and each ISI level. Small dots depict individual CEs and slopes values. CE mean values decrease as a function of the ISI, with exception of the ISI_{2000} condition of the <CS> group. However, CE mean values are higher for the <CS> than for the <SC> group. **(b)** RMSEs of Bayesian observers. Bayesian observer's responses successfully simulated the main CE results of the human observers: (1) the CEs decrease as the ISI increases, and (2) CEs values are higher for the <CS> than for the <SC> group.

replicated our finding that, by increasing the ISI between the first and second event, CEs decrease, although significantly less frequently for the <CS> group²⁴.

Classical TOE studies suggest that the level of noise in the internal representation of the 1st stimulus is larger than the 2nd stimulus' noise due to the encoding process and the maintenance of the 1st stimulus in memory^{27,28}. Likewise, recent results in the auditory system propose a Bayesian model where the discrimination of two stimuli is done by comparing the second tone versus the decaying average of the first tone¹⁴. Thus, errors in temporal discrimination should occur during memory retrieval and decision-making processes. Such a Bayesian model predicts that participants will have a better performance with a short ISI. However, our results show that increasing maintenance in memory does not have a detrimental effect on accuracy, on the contrary it has beneficial effects, possibly by reducing attentional blink effects. Indeed, our results are consistent with Grondin's findings showing that the CE is reduced when the duration of the ISI is 1.5 s⁹. Grondin found that this benefit occurs when using both single and multiple visual standard stimuli. We suggest that these results can be explained by the beneficial effects of allocating attention in time, oriented to the encoding of unpredictable events^{29–31}. Attention helps when S is in the first position, as it enhances the encoding of the C stimulus whose duration is unpredictable. We used this assumption to build our Bayesian model: allocating attention in time was modeled by decreasing the noise of the sensory input when the ISI increased.

The third finding concerns the preeminence of the ordinal position of the longer stimulus—independently of event type (S or C)—in driving accuracy. This offers a novel serial order bias in serial perception based on duration-dependent relative positions of stimuli, regardless of their duration relative to the global stimulus distribution. When the first stimulus in a series is shorter than the second stimulus, regardless of whether it is S or C, participants were biased to say that the first event was longer, consistently making mistakes. We found this effect in all experimental conditions of the <CS> group, but for the <SC> group it was only present at the shortest ISI (400 ms). We also showed that the serial order bias or perceptual glitch arises at all sensory levels of the <CS> group but only at the weak sensory level of the <SC>. These patterns of results explain why the magnitude of the time distortions (indexed via the CE) are larger for the <CS> than for the <SC> group. The serial order bias can be minimized by an increase of the ISI or an increase in the level of the sensory evidence if the first stimulus is predictable (<SC> group). However, when a dynamic, unpredictable stimulus is displayed in first position (<CS> group), such a bias is at ceiling and leading to temporal errors that will increase the magnitude of the time distortions across the board.

Fourth, our Bayesian model successfully simulated our human behavioral findings and showed that the Type A effect arises under sensory uncertainty because of the highlighted serial perceptual encoding inefficiency. Our modelling captures the idea that the temporal stimulus's perception in the visual system is shaped not only by sensory noise but also by a perceptual bias that systematically makes participants expand the first stimulus in a series if unpredictable. Naturally, if such stimulus is already longer than the second, the bias would not be visible as it would not lead to perceptual mistakes. By simulating our behavioral results with a Bayesian model, we provide more evidence to show that time estimation and duration discrimination are a dynamic process that not only take in consideration the current sensory information of the two stimuli but also the information of previous trials. Note that while we were not able to model the Type B effect, by modelling the Type A effect our findings align with recent results showing that the type A effect is best modeled with a Bayesian model using a Kalman filter¹².

Our behavioral results and simulations highlight the importance of the ISI factor in 2IFC paradigms. Both De Jong et al.¹² and Raviv et al.¹⁴ proposed powerful models for temporal discrimination, but implicit in their experimental paradigms and models is the employment of a long ISI. De Jong and colleagues implemented an ISI of 1000 ms, whereas Raviv et al., used an ISI of 950 ms. Here, we modeled the ISI's effects by decreasing the noise of the sensory input as the ISI increased. To do that, we implemented different distributions for the internal representation's noise.

The replication of our behavioral results with a Bayesian model offers an insight on the computations that the human brain might use as a strategy for temporal discrimination. At the same time, it shows how this computation is affected by sensory noise, the allocation of attention in time, and a hitherto unknown perceptual bias. As our stimuli used visual intervals at the bottom of the sub-second scale (120 ms), the question remains as to whether the novel perceptual serial order glitch we discovered would disappear for stimulus intervals in the supra-second range, and whether it is present in other sensory modalities, besides vision^{45,46}. Future research is needed to uncover the physiological basis of such a strong, implicit expectation about the temporal statistics of incoming stimuli which can drive humans to distort time perception under uncertainty.

Materials and methods

Participants. The experiment was organized as a between-subject design, with separate groups for the position of the stimuli: S in first position (<SC>) and C in first position (<CS>). Part of the results of the <SC> group were previously published²⁴. This dataset has a sample of 52 participants (34 female; ages: 18–33; mean age: 24.42). One participant was removed due to chance level accuracy (< 55%). Therefore, the final sample included the data from 51 participants (33 female; ages: 18–33; mean age: 24.45). For the <CS> group we had an initial sample of 58 participants (45 female; ages: 18–37; mean age: 25.41). Four participants were removed due to chance level accuracy (< 55%). Therefore, the final analysis included the data from 54 participants (41 female; ages: 18–33; mean age: 25.31). In total, we report on the behavior of 105 participants. For the analyses of the slope and constant error (CE), participants were excluded when one of the dependent variables had a value with three standard deviations above (or below) the mean. Thus, for the analyses of the slope and CE, nine participants were excluded following this procedure.

Individuals were recruited through online advertisements. Participants self-reported normal or corrected vision and had no history of neurological disorders. Up to three participants were tested simultaneously at computer workstations with identical configurations. They received 10 euros per hour for their participation.

Ethics statement. The studies were carried out in accordance with the the code of ethics of the World Medical Association (Declaration of Helsinki) for studies involving humans, and were approved by the Ethics Committee of the Max Planck Society. Written informed consent was obtained from each participant before starting the session.

Design. We used a classical interval discrimination task by implementing a 2IFC design, where participants were presented with two visual durations: S and C^{6,32}. S had a magnitude of 120 ms. For the <SC> group S was always displayed in the first position, but it was shifted to second position for the <CS> group. In both groups, we used three magnitudes for the step comparisons between S and C: 20, 60, and 100 ms. We derived the magnitudes for the C stimuli as $S \pm \Delta$, which resulted in the next C intervals: 20, 60, 100, 140, 180, and 220 ms. C stimuli were randomized on a trial-by-trial base.

We used the same four ISIs for both groups: 400, 800, 1600, and 2000 ms. For each trial, the inter-trial interval (ITI) was randomly chosen from a uniform distribution between 1 and 3 s. Participants judged whether the S or C stimulus was the longer duration. They responded by pressing one of two buttons on an RB-740 Cedrus Response Pad (<http://cedrus.com>) and were provided with immediate feedback on each trial.

Stimuli and apparatus. Stimulus duration was determined as a succession of two blue disks with a diameter of 1.5° presented on a gray screen³³. Empty stimuli were implemented to ensure that participants were focused on the stimuli's temporal properties³⁴. All stimuli were created in MATLAB R2018b (<http://mathworks.com>), using the Psychophysics Toolbox extensions^{35–37}. Stimuli were displayed on an ASUS monitor (model: VG248QE; resolution: 1920×1080; refresh rate: 144 Hz; size: 24 in) at a viewing distance of 60 cm. When the C intervals (20, 60, 100, 140, 180, and 220 ms) were transformed into video frames and resulted with decimal fractions, they were rounded to the nearest integer. Thus, the C intervals had 3, 9, 15, 21, 26, and 32 frames, respectively.

Protocol (Task). The experiment was run in a single session of 70 min. Participants completed a practice set of four blocks (18 trials in each block). All sessions consisted of the presentation of one block for each ISI condition. Each block was composed of 120 trials. For each ISI the C intervals were presented in random order. Each ISI block was also presented in random order, each randomization was unique.

To avoid fatigue, participants always had a break after 60 trials. Each trial began with a black fixation cross (diameter: 0.1°) displayed in the center of a gray screen. Its duration was randomly selected from a distribution between 400 and 800 ms. After a blank interval of 500 ms, S was displayed and followed by an ISI. After this, C was displayed. Participants were instructed to compare the interval of the two stimuli by pressing the key “left”, if S was perceived to have lasted longer, and the key “right” if C was perceived to have lasted longer. After responding, they were provided with immediate feedback: the fixation cross color changed to green when the response was correct, and to red when the response was incorrect.

Data analysis. Data cleansing was implemented with Python 3.7 (<http://python.org>) using the ecosystem SciPy (<http://scipy.org>). GLMMs were fitted in R³⁸ using the *lme4* package (version 1.1.21). Estimated marginal means (EMM) were computed for the significant interactions in the fitted model to perform post-hoc comparisons and contrasts using the *emmeans* package³⁹ (version 1.7.3). Bonferroni adjustment of *p*-values for multiple comparisons was applied where applicable. Raincloud plots were created using the *raincloud* function for R⁴⁰. All data analyses and simulations, whether using Python or R, were performed in Jupyter Lab (<http://jupyter.org>). The annotated notebooks can be consulted at Open Science Framework (OSF) (<https://osf.io/qnj3t/>).

General Linear Mixed Models (GLMMs). We modeled our behavioral data with GLMMs to estimate a single model across all subjects and distinguish within- and between-subjects errors^{10,25,41}. To fit the GLMMs we input the responses as a whole^{10,12}. Raw accuracy was analyzed with GLMMs using a binomial parameter with a logit link function. To compute the temporal sensitivity and the PSE we fitted GLMMs using as dependent variable the percentage of responses “C > S”. We used again a binomial parameter but this time using a probit function. We calculated the expected value of the responses as follow:

$$\phi^{-1}[P(Y_{ij} = 1)] = \beta_0 + \beta_1 x_{ij} \quad (1)$$

where x_{ij} is the Comparison stimulus’ duration, Y_{ij} the response variable for subject I and trial j . If the C stimulus is judge longer than S, then $Y_{ij} = 1$; but $Y_{ij} = 0$, if C is shorter than S. The probability of the response “C longer than S” $P(Y_{ij} = 1)$ is linked with the linear predictor via the probit link function ϕ^{-1} . The fixed-effect parameters β_0 and β_1 are the intercept and the slope, respectively. The β_1 is an index of the temporal precision, which is also called the Just Noticeable-Difference (JND)¹⁰. The PSE is a function of both parameters:

$$PSE = -\frac{\beta_0}{\beta_1} \quad (2)$$

We derived the CE as the difference between the PSE and the magnitude ϕ_s of the Standard duration: $CE = PSE - \phi_s$, and $CE = \phi_s - PSE$, for the <SC> and <CS> groups, respectively.

To apply Model Comparison (BMC) to the GLMMs and decide between models, we applied the Akaike Information Criterion (AIC), whereby the model with the lowest AIC was selected as the one that best explained the data⁴².

Bayesian modelling. We implemented a Bayesian model to replicate our results by using a Kalman filter. We based our Bayesian model on the work of Petzschner and Glasauer²², Glasauer and Shi²³, and de Jong et al.¹². In this model the prior represents the intervals stored in memory (that is, the internal representation of previous trials), the likelihood is the current sensory input, and the posterior is the current estimate or *percept*. To run our Bayesian model and use it for temporal discrimination, we used the duration of both stimuli—S and C—as inputs for this model. To do that we used the representation of the stimulus’ duration on a logarithmic scale and added some Gaussian noise:

$$x_m = \ln(d) + n_m \quad (3)$$

where d is the physical duration of the stimulus (S or C) in a linear scale and x_m is the internal noisy representation. The random variable n_m represents the normally distributed measurement noise $p(n_m) \approx N(0, \sigma_m^2)$ ⁴³. To run simulations for individual participants we randomly selected values for σ_m^2 from a truncated normal distribution. Note that the magnitude of σ_m^2 is the temporal sensitivity of each participant. The model compared on a trial basis the logarithmic representations of S and C, and yielded 0 or 1:

- the model yielded 0 if “S > C”;
- the model yielded 1 if “C > S”;

when a new duration—indexed by n —is perceived it is represented by the likelihood function, which is a Gaussian distribution with $p(x_{m,n}) \approx N(x_{m,n}, \sigma_m^2)$. Note that S and C are the mean of the priors that emerged from perceiving both stimuli. The prior too is modelled as a Gaussian distribution: $N(\mu_p, \sigma_p^2)$. To estimate a stimulus’ duration the prior is updated through a weighted average of the previous prior distribution and the currently sensed likelihood. For each measurement the update step is modeled by the formulation of the Kalman filter for a 1D first-order system:

$$k_n = \frac{p_{n-1} + q}{p_{n-1} + q + r} \quad (4)$$

$$p_n = k_n \cdot r \quad (5)$$

where r is the uncertainty of the current likelihood (σ_m^2) and p_{n-1} the uncertainty of the previous prior ($\sigma_{p,n-1}^2$). Thus, the Kalman gain (k) of a new observation is determined by both uncertainties and a process variance q . The variance of the prior system p_n is updated by the product of the Kalman gain and r , whereas the prior mean $\mu_{p,n}$ is updated as follow:

$$\mu_{p,n} = (1 - k_n) * \mu_{p,n-1} + k_n * x_{m,n} \quad (6)$$

To simulate our behavioral results, we used three assumptions for modelling the noise n_m associated with the internal representation x_m . (1) We assumed that n_m decreases as the duration of the ISI increases. That is, the noise n_m associated with the second stimulus was larger for the ISI₄₀₀ than for the ISI₂₀₀₀ condition. (2) We assumed that independently of the group (<SC> or <CS>) or even type (S or C), the stimulus' noise come from two different distributions. Thus, we used a distribution for the "Longer stimulus in 1st position" stimuli' noise and a second one for the "longer stimulus in 2nd position" stimuli' noise. (3) Because the perceptual bias is stronger at the weak sensory level than the other two sensory levels, we used a truncated normal distribution for n_m at the weak sensory level ($\Delta 20$). For the <SC> group this bias affected the (+ $\Delta 20$) trails, but for the <CS> group affected the $-\Delta 20$ trials. See the annotated notebooks at OSF (<https://osf.io/qnj3t/>).

For each group we generated data for 100 participants with 120 trials for each ISI level. For both groups we kept a constant value for q ($q = 1.5$).

Root Mean Squared Error (RMSE). To compare the performance of human participants against the Bayesian observer's responses, we computed the RMSE which is given by the standard deviation (SD) and the bias: $RMSE^2 = SD^2 + bias^2$, where the SD is the slope and the bias is the CE^{19,44}. As the RMSE can be written as the standard equation of the circle, it provides an effective geometric and graphical description to track changes on the CE and explore the tradeoff between the CE and the temporal precision. RMSEs are given by the distance of the origin, and are depicted by a quarter circle. Because we had negative values for the CE, we took -60 as the origin point instead of 0. Thus, to find the circle's intercept on the x -axis, that is the axis of the CE, we took the absolute distance between -60 and the CE.

Data availability

Anonymized data are available at the Open Science Framework (<https://osf.io/qnj3t/>).

Received: 21 May 2022; Accepted: 1 December 2022

Published online: 06 December 2022

References

- Jazayeri, M. & Shadlen, M. N. A neural mechanism for sensing and reproducing a time interval. *Curr. Biol.* **25**(20), 2599–2609 (2015).
- Nakajima, Y., Hoopen, G. T., Hilkhuisen, G. & Sasaki, T. Time-shrinking: A discontinuity in the perception of auditory temporal patterns. *Percept. Psychophys.* **51**(5), 504–507 (1992).
- Merchant, H. & De Lafuente, V. Introduction to the neurobiology of interval timing. *Neurobiol. Interval Timing* **2014**, 1–13 (2014).
- Schab, F. R. & Crowder, R. G. The role of succession in temporal cognition: Is the time-order error a recency effect of memory?. *Percept. Psychophys.* **44**(3), 233–242 (1988).
- Fechner, G. T. *Elemente der Psychophysik*, volume 2 (Breitkopf u. Hartel, 1860).
- Wright, B. A., Buonomano, D. V., Mahncke, H. W. & Merzenich, M. M. Learning and generalization of auditory temporal-interval discrimination in humans. *J. Neurosci.* **17**(10), 3956–3963 (1997).
- Buonomano, D. V. & Maass, W. State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Rev. Neurosci.* **10**(2), 113–125 (2009).
- Grondin, S. Methods for studying psychological time. *Psychol. Time* **51–74**, 2008 (2008).
- Grondin, S. Discriminating time intervals presented in sequences marked by visual signals. *Percept. Psychophys.* **63**(7), 1214–1228 (2001).
- Moscattelli, A., Mezzetti, M. & Lacquaniti, F. Modeling psychophysical data at the population-level: The generalized linear mixed model. *J. Vis.* **12**(11), 26–26 (2012).
- Lapid, E., Ulrich, R. & Rammsayer, T. On estimating the difference limen in duration discrimination tasks: A comparison of the 2afc and the reminder task. *Percept. Psychophys.* **70**(2), 291–305 (2008).
- de Jong, J., Akyurek, E. G. & van Rijn, H. A common dynamic prior for time in duration discrimination. *Psychon. Bull. Rev.* **28**(4), 1183–1190 (2021).
- Ashourian, P. & Loewenstein, Y. Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS ONE* **6**(5), e19551 (2011).
- Raviv, O., Ahissar, M. & Loewenstein, Y. How recent history affects perception: The normative approach and its heuristic approximation. *PLoS Comput. Biol.* **8**, 10 (2012).
- Alcalá-Quintana, R. & García-Pérez, M. A. A model for the time-order error in contrast discrimination. *Q. J. Exp. Psychol.* **64**(6), 1221–1248 (2011).
- Hellström, Å. Factors producing and factors not producing time errors: An experiment with loudness comparisons. *Percept. Psychophys.* **23**(5), 433–444 (1978).
- Hellström, Å. Time errors and differential sensation weighting. *J. Exp. Psychol. Hum. Percept. Perform.* **5**(3), 460 (1979).
- Dyjas, O., Bausenhardt, K. M. & Ulrich, R. Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Atten. Percept. Psychophys.* **74**(8), 1819–1841 (2012).
- Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat. Neurosci.* **13**(8), 1020–1026 (2010).

20. Acerbi, L., Wolpert, D. M. & Vijayakumar, S. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput. Biol.* **8**(11), e1002771 (2012).
21. Shi, Z., Church, R. M. & Meck, W. H. Bayesian optimization of time perception. *Trends Cogn. Sci.* **17**(11), 556–564 (2013).
22. Petzschner, F. H. & Glasauer, S. Iterative bayesian estimation as an explanation for range and regression effects: A study on human path integration. *J. Neurosci.* **31**(47), 17220–17229 (2011).
23. Glasauer, S. & Shi, Z. The origin of Vierordt's law: The experimental protocol matters. *PsyCh. J.* **10**(5), 732–741 (2021).
24. Sierra, F., Poeppel, D. & Tavano, A. Two attentive strategies reducing subjective distortions in serial duration perception. *PLoS ONE* **17**(3), e0265415 (2022).
25. Agresti, A. *Categorical Data Analysis* (Wiley, 2002).
26. Allan, L. G. The perception of time. *Percept. Psychophys.* **26**(5), 340–354 (1979).
27. Bull, A. R. & Cuddy, L. L. Recognition memory for pitch of fixed and roving stimulus tones. *Percept. Psychophys.* **11**(1), 105–109 (1972).
28. Wickelgren, W. A. Associative strength theory of recognition memory for pitch. *J. Math. Psychol.* **6**(1), 13–61 (1969).
29. Nobre, A. C., Correa, A. & Coull, J. T. The hazards of time. *Curr. Opin. Neurobiol.* **17**(4), 465–470 (2007).
30. Rohenkohl, G., Cravo, A. M., Wyart, V. & Nobre, A. C. Temporal expectation improves the quality of sensory information. *J. Neurosci.* **32**(24), 8424–8428 (2012).
31. Cravo, A. M., Rohenkohl, G., Wyart, V. & Nobre, A. C. Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *J. Neurosci.* **33**(9), 4002–4010 (2013).
32. Mauk, M. D. & Buonomano, D. V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
33. Kliegl, K. M. & Huckauf, A. Perceived duration decreases with increasing eccentricity. *Acta Physiol. (Oxf.)* **150**, 136–145 (2014).
34. Grondin, S., Meilleur-Wells, G., Ouellette, C. & Macar, F. Sensory effects on judgments of short time-intervals. *Psychol. Res.* **61**(4), 261–268 (1998).
35. Brainard, D. The psychophysics toolbox. *Spat. Vis.* **10**(4), 433–436 (1997).
36. Pelli, D. G. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
37. Kleiner, M., Brainard, D. & Pelli, D. What's new in psychtoolbox-3?. *Perception* **36**, 1–16 (2007).
38. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018).
39. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. Emmeans: Estimated marginal means, aka least-squares means. *R Pack. Vers.* **1**, 2021 (2018).
40. Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R. & Kievit, R. A. Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Res.* **4**, 2356 (2019).
41. Cox, D. R. The regression analysis of binary sequences. *J. R. Stat. Soc.: Ser. B (Methodol.)* **20**(2), 215–232 (1958).
42. Yosiyuki, S., Ishiguro, M. & Kitagawa, G. Akaike information criterion statistics. *Dordrecht The Netherlands D. Reidel* **81**, 26853 (1986).
43. Stocker, A. A. & Simoncelli, E. P. Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* **9**(4), 578–585 (2006).
44. Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M. & Burr, D. C. Optimal encoding of interval timing in expert percussionists. *J. Neurosci.* **32**(3), 1056–1060 (2012).
45. van Wassenhove, V., Buonomano, D. V., Shimojo, S. & Shams, L. Distortions of subjective time perception within and across senses. *PLoS ONE* **3**(1), e1437 (2008).
46. Merchant, H. & de Lafuente, V. Introduction to the neurobiology of interval timing. *Neurobiol. Interval Timing* **1–13**, 2014 (2014).

Acknowledgements

We thank Luigi Acerbi, Federico Adolphi, Christina Lubinus, Kanthida van Welzen, Emily Ip, and Piermatteo Morucci for critical comments and suggestions.

Author contributions

F.S. and A.T. conceived the experiment(s), F.S. conducted the experiment(s), F.S., A.T. and R.M. analysed the results. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022